DOCUMENT RESUME

ED 422 887 IR 019 028

AUTHOR	Wisher, Robert A.; Curnow, Christina K.
TITLE	Techniques for Evaluating Distance Learning Events.
PUB DATE	1998-00-00
NOTE	7p.; In: Distance Learning '98. Proceedings of the Annual Conference on Distance Teaching & Learning (14th, Madison, WI, August 5-7, 1998); see IR 018 976.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Course Evaluation; *Distance Education; Evaluation Criteria; *Evaluation Methods; Evaluation Research; Feedback; Institutes (Training Programs); *Instructional Effectiveness; *Military Training; Participant Satisfaction; Program Effectiveness; Questionnaires
IDENTIFIERS	Army National Guard; National Guard

ABSTRACT

This paper describes techniques for creating a simplified form for evaluating the effectiveness of a distance learning event (i.e., a training or educational program occurring within 1 day) for which the development of comprehensive measures is often not feasible because the brevity of the event precludes the justification for a lengthy evaluation. The paper focuses on distance learning events for the National Guard Bureau. Background is provided on evaluation measures, including student reaction, learning, behavioral criteria, and results criteria. Development of the compressed evaluation form--one page of five assessment categories (course, technology, instructor, demographics, and motivation) with a total of 22 items--is discussed. Testing of the compressed evaluation form is described. Eight distance learning events were sampled, representing the primary distance learning events available on a nation-wide basis during the evaluation time frame. Training was delivered through a multi-point, one-way video, two-way audio connection through satellite links. A total of 1,306 soldiers participated in the training; the mean rate of returned evaluations was 74%. Two tables present data on student ratings of events and cross tabulation of amount learned by previous training. It is concluded that a wealth of information can be obtained from a simple, one-page evaluation form. (DLS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Techniques for Evaluating Distance Learning Events

Robert A. Wisher
Senior Research Psychologist
U.S. Army Research Institute

Christina K. Curnow
Research Fellow
George Washington University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C.H. Olgren
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 422 887

Introduction

The effectiveness of distance learning, when compared to traditional education and training settings, has been demonstrated hundreds of times (Russell, 1996). The studies reported, however, have been largely oriented to college courses, continuing education credits, or professional development training that occur over extended periods and which have "built-in" evaluation measures, such as a final exam. There has been little technical work in developing evaluation measures for short-term training events. This paper describes techniques for creating a simplified form for evaluating the effectiveness of a distance learning event. Here, a distance learning *event* refers to a training or educational program occurring within *one day*. For such abbreviated training the development of comprehensive measures is often not feasible as the brevity of the event precludes the justification for a lengthy evaluation.

The sponsor for this work is the National Guard Bureau, which plans to connect classrooms in 54 states and territories as one distance learning network. While regular courses are being converted to the distance learning format, most of the training conducted during the early stages of the transition was oriented to the abbreviated training event. Thus the need for the specialized evaluation form to verify student learning. Studies of distance learning in the Army have demonstrated positive results. For example, in a study that examined the cost effectiveness of audio teletraining for unit clerks in the National Guard, considerable cost savings over the residence training were documented for the distance learning format (Wisher, et. al., 1997).

Background

Early evaluation studies in distance learning were mostly descriptive case studies that focused on learner satisfaction, and were often anecdotal (OTA, 1989). More recently, there have been discussions regarding what methods are best to use for evaluating distance learning programs. In their review of the evaluation literature, Harrison et. al. (1991) identified three unique components of distance learning that consistently emerged: instruction, management, and logistics (i.e., technology factors, technical qualities, environment, etc.). From this point of view, the sponsor was primarily interested in the instructional effectiveness and technology factors.

As with evaluations of other modes of training, the instructional effectiveness of a distance-learning training event can be measured in terms of student reaction, learning, behavioral criteria, or results criteria. Each measure is used to assess different aspects of the value of an event. Some are objective and some are subjective, and each has advantages and

IR019028

disadvantages in terms of preparation of instruments, administration time, and decisiveness of results.

An evaluation based on student reactions measures how favorably the participants respond to a training event, including its contents and delivery, relevance to their job, the training technologies used, and the overall training environment. Favorable reactions to training events do not guarantee that learning has taken place, but they are useful to collect for several reasons. First, positive reactions help gain or maintain organizational support for training events while negative reactions can lead to problems of future support. The second benefit is that reaction measures can serve as a source of immediate feedback to the training providers, including instructors, production staff, and training event organizers. Such data can prove helpful in planning and designing future training events. A third benefit, not necessarily unique to reaction measures, is subgroup analysis, through cross tabulation. Such an analysis can further elucidate the relative impact of the training across subgroups, such as military rank.

Another type of reaction measure, obviously subjective, is the self-assessment variable. This technique requires that respondents evaluate themselves on various dimensions of the training, such as how much more they learned compared to what they already knew about the topic. The self-assessment variable may be viewed as a bridge between a reaction measure and a learning measure, as it assesses learning directly, but in a subjective way. Regardless of the context, most of the research in this area has identified significant correlations between self-assessments of ability to perform tasks and performance measures (either knowledge tests or supervisor assessment).

In contrast to the subjective nature of reaction measures, learning criteria offer an objective means to assess the knowledge and skills acquired during the training program. Although learning criteria are a stronger, more decisive measure, one drawback concerns the time to develop test instruments and the costs of their administration—the evaluation resources. Tests specific to the knowledge taught need to be developed, pre-tested, and administered. A learning measure usually requires additional development effort which might be impractical for a short-term event. The use of self-assessment as a surrogate learning measure, although subjective, might remedy this issue.

Development of the Compressed Evaluation Form

In the context of distance learning training events for the National Guard Bureau, the reaction and learning measures were most relevant due to the need for quick feedback to event organizers and sponsors as well as to monitor whether the training objectives of the events are being met. One of the practical considerations was the effort required for the development of a knowledge or performance test for the learning measure. In view of the limited resources for instrument development and the range of anticipated training events, a strategy to have a self-assessment scale serve as the learning measure was selected. This set the stage for a compressed evaluation form to be created and the subsequent evaluation time to be brief, since a lengthy knowledge test would not be included. More importantly, it provided an acceptable source of data for the types of analyses relevant to the Bureau's interests, mainly the instructional and technology factors: How well did the technology work and did the soldiers learn from the training experience?

Recommendations from the literature on length, saliency, confidentiality and anonymity, and ease of return were factored into the design of the evaluation form. Additional suggestions were gained from other evaluators in the distance learning field. A review of instruments from educational institutions and government agencies, most published in the open literature but some made available through personal communications, resulted in the creation of a set of comparative variables: number of pages for an instrument, number of assessment categories (such as demographic, instructor, facilities etc.), number of questions within each category, and a description of the scales used. The results of this analysis showed the following: the length of the evaluation instrument ranged from 1 to 5 pages with an average of 37 items covering five assessment categories and using categorical and 5-point Likert scales. Based on this analysis and the Bureau's requirement ("did the technology work and did soldier's learn?"), the evaluation form was compressed into a single page of five assessment categories (course, technology, instructor, demographics, and motivation) with a total of 22 items. The demographic variables were needed to detail a more refined look at learning patterns that might emerge from a cross tabulation of the data. The goals of a short form with a confidentiality statement, salient to-the-point items, and a pre-paid return envelope provided to each site were accomplished.

Test of the Compressed Evaluation Form

Eight distance learning events were sampled. These events represented the primary distance learning events available on a nation-wide basis during the evaluation timeframe. All training was delivered through a multi-point, one-way video two-way audio connection through satellite links. Examples of the training events were Risk Management (three iterations), Airborne Call For Fire, and Terrorism Update. A total of 1,306 soldiers participated in the training. The number of remote sites per event ranged from 3 to 32, with up to 63 students per site. Site facilitators returned attendance lists to an event coordinator, allowing return rates to be computed. The percentage of returned evaluation forms ranged from 30% to 97%, with a mean return rate of 74%.

Nine questions addressed technology factors, the course, the instructor, and the learning environment. Each question asked respondents to rate an aspect of the event using a five-point scale from "poor" (1) to "excellent" (5). The means and standard deviations for each appear in Table 1. Also presented are means from substantially similar questions reported in studies of distance learning conducted by the Navy (Wetzel et .al., 1996). These provide a benchmark comparison for the ratings obtained in the present study. By and large the students reacted favorably to the experience in agreement with the Navy study. The lowest mean ratings occurred for "opportunity to ask questions" and "quality of audio" and the highest ratings were for "location of video" and "quality of video."

A self-assessment measure was used as a surrogate for a learning measure. The form of self-assessment in the current report stems from the question "Compared to what you already knew about *'course topic,'* how much more did you learn in this training event?" Since resources for the evaluation were limited and the variety of training events was considerable, the development of more decisive measures of learning was not practical, nor were they already available from end-of-course tests.

Table 1. Means and Standard Deviations for Student Ratings of Events

Question	Navy Benchmark	National Guard Events		
		M	*SD*	*n*
Location of the video screen	4.6	4.4	.81	1031
Quality of audio	4.1	3.8	1.21	1023
Quality of video	4.5	4.2	.96	1025
Instructor effectiveness	—	3.9	.91	960
Opportunity to ask questions	—	3.7	1.22	903
Responsiveness to student questions	4.5	3.9	1.00	849
Relevance of course to guard duties	4.3	4.1	.96	938
Overall learning environment	—	4.0	.99	1005
Overall effectiveness of instruction	—	3.8	.98	1000

Cross Tabulations

Overall, 58% of the respondents reported previous training in the topic being covered. Of particular interest were the interrelationship between two variables, the self-assessment of perceived learning and whether or not the respondent had previous training on the topic. This interrelationship was examined by using cross-tabulation, resulting in Table 2. Note that on the five point scale, ratings of 1 or 2 were netted as "little" was learned, a rating of 3 was interpreted as "some" was learned, and a rating of 4 or 5 was netted as "a lot" was learned. In all cases, the ratings were made relative to whether the respondents had previous training on the topic.

Table 2. Percentage of Responses for the Cross Tabulation of Amount Learned by Previous Training

	No Previous Course			Previous Course			
	Amount Learned			Amount Learned			
	Little	Some	A lot	Little	Some	A lot	*n*
Overall (all 8 courses)	12%	32%	56%	26%	43%	32%	1,044
Risk Mgt. Course							
Iteration 1	6%	32%	62%	22%	46%	32%	227
Iteration 2	13%	35%	52%	17%	50%	33%	88
Iteration 3	13%	27%	60%	28%	41%	31%	234

Table 2 shows a pattern in which those who had not taken a previous course in the topic area reported greater amounts of learning. Overall, more than half (56%) of those with no previous course reported learning "a lot." However, for those reporting having taken a previous course (and thus possessing prior knowledge) only 32% reported learning "a lot." While the majority of respondents reported learning "some or a lot," it appears that the courses were generally geared to those individuals who had no previous training in the topic area. In comparison, many more of the respondents (26% vs. 12%) who had previous training learned "little," indicating that a more advanced course may have been more appropriate for them. This supports the face validity of the self-assessment technique.

As the Risk Management course was evaluated on three separate occasions, it presented the best opportunity to explore the consistency of the relationship between the two key learning variables. As can be seen in Figure 2, the patterns between the three courses were quite similar. Face validity was examined by comparing perceived learning between subgroups who either had or did not have a previous course on the topic. Analyses showed significant differences in self-assessed learning between those who had taken a previous course and those who had not ($F(1,476) = 28.16$, $p < .001$). To gauge the external reliability of the previous course experience variable and the self-assessed learning variable, a 2 X 3 ANOVA was conducted, with two levels of previous course experience and three levels of learning. This analysis revealed no significant mean differences of self-assessed learning among the three events ($F(2, 476) = .38$, $p = .684$). The ANOVA results indicate that the amount of perceived learning was consistent across all three events. Furthermore, there were consistent differences in perceived learning between those who had and had not taken a previous course in Risk Management.

Conclusions

A useful strategy for evaluating single-day distance learning events is the use of a compressed, one-page evaluation form that is designed with consideration of confidentiality, saliency, convenience of return, and length. The research literature has demonstrated these factors to be of value in obtaining higher return rates. Another aspect of the literature that was useful was the self-assessment variable for evaluating learning. Even though this variable has had some controversy in the past, its use in a military training setting (especially when coupled with anonymity) appears not to have the same problems that have troubled other applications. The obvious advantage in the self-assessment approach is a great savings in administration time, not to mention the avoidance of having to develop a separate learning evaluation instrument.

The findings and analyses reported above demonstrate the wealth of informative data that can be obtained from a simple, one-page evaluation form. There are many other cross-tabulations and analyses that could be conducted, depending on the interests of the organizations, to pinpoint technical shortfalls, course effectiveness, learning, and certain policy issues. Although the compressed form is only one page in length, the opportunities to cross tabulate on the basis of demographic factors can lead to important insights and trends. Such analyses can provide useful feedback to the stakeholders in distance learning, including organizers, managers, instructors, and technicians.

References

Harrison, P., Seeman, B., & Behm, R. (1991) Development of a distance education assessment instrument, *Educational Technology Research and Development, 39,* 65–77.

OTA (1989). *"Linking for learning" A new course for education.* U.S. Congress Office of Technology Assessment (OTA-SET-430). Washington, DC: U.S. Government Printing Office.

Russell, T.L. (1996). *The "no significant difference" phenomenon.* [Online] Available: http://tenb.mta.ca/pheno/phenom.html.

Wetzel, D., Radtke, P., Parchman, S., & Seymour, G. (1996). *Delivery of a fiber optic cable repair course by videoteletraining* (NPRDC-TR-96-4). San Diego, CA: Navy Personnel Research and Development Center.

Wisher, R. A., Priest, A. N. & Glover, E. (1997) *Audio teletraining for unit clerks: a cost-effectiveness analysis.* (Research Report 1712) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Autobiographical Sketches

Dr. Robert A. Wisher is a Senior Research Psychologist with the U.S. Army Research Institute in Alexandria, Va. He is responsible for evaluating the effectiveness of distance learning in the Army National Guard and for maintaining a technology watch for emerging training technologies that could be applied to Army training requirements. He holds a Ph.D. in experimental psychology from the University of California, San Diego.

Address: U.S. Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333-5600
Email: Wisher@ari.army.mil
Phone: (703) 617-5540
Fax: (703) 617-3573

Ms. Christina K. Curnow is a Research Fellow with the U.S. Army Research Institute, assigned to the Advanced Training Methods Research Unit. Her research responsibilities center around designing, developing, and analyzing distance learning evaluations for the Army National Guard. Additionally, she conducts field research on self assessment accuracy of long-term memory for trained skills and knowledge. She is currently completing her doctoral studies in Industrial and Organizational Psychology at George Washington University.

Email: Curnow@ari.army.mil